



## 특집 07

# 차량용 다국어 음성 인터페이스 기술



송민규 (미디어젠)

- 목 차 »
1. 서 론
  2. 차량용 음성 기술 동향
  3. 다국어 음성 미들웨어와 언어 현지화
  4. 차량용 음성 미들웨어 발전 방향
  5. 결 론

## 1. 서 론

현대인의 생활에서 자동차는 단순한 운송수단으로만 그치는 것이 아니라, 하나의 새로운 생활 공간으로 자리 잡고 있다. 자동차 안에서 음악과 각종 콘텐츠, 애플리케이션 등을 이용할 수 있는 스마트카의 개발이 국내 자동차 제조사를 비롯한 세계 유력 완성차 업체들을 중심으로 빠르게 진행되고 있다. 스마트카란 인포테인먼트와 텔레매틱스 기능을 겸비한 자동차를 의미하는데<sup>[1]</sup>, 음성 인식 기술은 다양한 기능과 콘텐츠를 제공할 수 있는 안전하고 편리한 전자 기기 제어 및 정보 검색의 인터페이스를 담당하는 기술로 큰 관심을 모으고 있다.

스마트카에서 음성 기술이 반드시 필요한 이유는, 방대한 양의 콘텐츠를 차량 내에서 영위할 수 있게 되었기 때문이다. 또한 운전 상황에서 터치나 필기 인식은 사고 위험을 내재하고 있으므로, 그보다 훨씬 안전한 음성 인터페이스 기술이 차

량 내 안전성 면에서 훨씬 적합한 기술로 평가받고 있다.

최근 모바일 시장을 중심으로 적용되어 큰 관심을 모으고 있는 서버 기반 음성인식 기술, 특히 딥러닝 기술까지 차량에 적용되면 웹 검색은 물론, 음성 E-Mail, 음성 SMS, 음성 소셜 네트워크(SNS) 및 소셜 커머스의 이용 등 무궁무진한



(그림 1) 미래형 스마트카

서비스를 차량 안에서 제공할 수 있게 된다. 차량용 음성 미들웨어는 이러한 차량 내 음성 인터페이스를 가능하게 하는 여러 요소 중의 하나이다.

본고에서는 양산 적용 사례를 중심으로 차량용 음성 미들웨어 및 다국어 적용을 위한 언어 현지화 방안에 대해 살펴보고, 향후 발전 방향에 대해 조망해 보고자 한다.

## 2. 차량용 음성 기술 동향

차량 IT 융합 기술은 첨단 전기 전자 제어 기술과 접목되어 안정성과 주행 효율성을 높이는 지능형 자동차의 단계를 넘어서 정보 커뮤니케이션 접근성을 높여주는 스마트카의 단계로 나아가고 있다.<sup>[1,2]</sup> 음성인터페이스 기술은 이러한 흐름에 발맞추어 가장 빠르게 진화하고 있는 차량 IT 융합 기술 중의 하나이다.

음성인터페이스 기술은 음성인식의 완성도와 음성합성 출력음의 자연성이 확보되면서 이를 이용한 응용 제품 및 다양한 서비스가 확산될 것으로 전망된다. 이동통신 단말기부터 로봇, 가전, 교육, 의료, 자동 통역, 홈 네트워크 등 지능형 음성 기술의 적용 분야는 다양한 산업에 걸쳐 성장 동력으로 영향을 미칠 수 있다.<sup>[3]</sup> 그 중에서도 텔레매틱스 분야는 안전성과 편의성이라는 큰 측면에서 음성 기술이 적용되기에 가장 적합한 분야라고 할 수 있다.

음성인식 기술은 크게 5단계로 세대 구분을 할 수 있는데, 현재 5세대로 분류되는 2007년 이후부터는 정보서비스를 처리하기 위한 대화 시스템과 유비쿼터스 환경에서의 서비스 응용이 가능할 정도의 성능 향상이 이루어졌다고 할 수 있다. 특히 차량 환경에서 양산 품질을 확보하기 위한 성능 수준을 만족하기 시작한 시기 또한 이 즈음이므로 5세대야말로 본격적인 양산 적용 세대에

속한다고 할 수 있다. 음성합성 기술은 자동차 운전 상황처럼 주의 분산을 방지해야 하는 특정한 환경에서 효율적으로 정보를 제공할 수 있는 수단으로서 높은 관심을 끌고 있다. 차량 내 정보 안내 및 시나리오 전개를 위해 적용되고 있는 음성 대화 시스템은 음성인식과 음성합성 엔진의 복합적 연동을 통해 구현된다.<sup>[4-6]</sup>

이처럼 음성 엔진의 성능이 양산 수준에 도달하면서, 사용자의 관심은 성능의 육성을 위해 도외시했던 사용성의 개선으로 옮겨가고 있다. 차량 내 음성 서비스의 사용성에 대한 연구는 여타 HMI 평가 기준과 동일하게 기억 용이성(Easy to Remember), 습득 용이성(Easy to Learn), 사용 효율성(Efficient to Use), 안전성(Safe to Use), 오락성(Enjoyable to Use), 효과성(Effective to Use) 등 다양한 측면에서 진행되고 있다.<sup>[7]</sup>

<표 1>은 OEM 및 After market 차량용 텔레매틱스 단말기 주요 업체의 음성인식 현황을 나타낸 것이다. 기존의 멀티미디어 및 내비게이션의 일반 기능을 음성으로 작동하는 단계를 넘어서서, 언어 처리 강화를 통한 문장형 인식이나 연결성(Connectivity)을 이용한 어플리케이션 연동 등 서비스를 강화하는 추세로 발전하고 있다.

차량용 음성 미들웨어는 이러한 추세에 동참하여 진화된 음성 시스템을 개발할 수 있는 기반을 제공할 수 있다. 음성인식 및 합성 엔진 제어부와 장치 어플리케이션 연동부의 변경을 최소화하면서, 시나리오 강화, 유사어 보강, 문장형 통계 데이터 축적, 다국어 현지화 강화 등 콘텐츠 부분의 경쟁력 강화를 도모할 수 있기 때문이다.

<표 2>와 같이 차량용 음성인식 시장의 동향은 크게 엔진 업체의 동향, 차량용 음성 산업 동향, 그리고 가치의 변화라는 관점에서 조망할 수 있다.

음성인식 엔진 분야는 이미 다국적 거대 기업 및 전문 기업으로 주도권이 완전히 넘어간 상태

〈표 1〉 주요 텔레매틱스 단말기 음성인식 적용 현황

주요 업체	모델	주요 기능
현대기아자동차	AVN / Audio	16개 다국어 음성인식 지원, 전화걸기 기능 강화
	Bluelink / UVO	센터 음성인식을 통한 다양한 서비스 제공
Ford(OEM)	Sync	“나 배고파”라고 말하면 주변 음식 리스트 제시 (POI 한정) 이용 중인 콘텐츠 분석하여 유사 콘텐츠 제공, “비슷한 음악 틀어”
	Sync Applink	스마트폰과 Sync를 블루투스로 연결하여 스마트폰 어플리케이션을 차량 음성인식으로 제어 (일부 기능)
Toyota(OEM)	Entune	스마트폰과 차량을 블루투스로 연결하여 각종 어플리케이션을 차량 음성인식으로 제어 (일부 기능)
BMW(OEM)	iDRIVE	PTT(Push to Talk) 버튼을 스티어링 휠과 하단 조작부에 각각 배치하여 음성인식 진입 용이
Jeep(OEM)	Uconnect	튜토리얼 모드에서 음성인식 시스템 전반에 대한 안내를 제공
TomTom	After market Navi	동일 명령에 대한 다양한 이형 표현 및 문장형 표현
Garmin	After market Navi	음성인식 진입 시 PTT 버튼 대신 특정 키워드를 인식하여 진입 (Softkey 방식 도입)

〈표 2〉 차량용 음성인식 시장 동향

구분	내용	설명
Player	규모의 경쟁	구글, 애플, MS 등 거대 다국적 기업의 자체 ASR 강화
	전문 기업 독점화	Nuance 엔진 개발 기업 독점적 지위 강화
Automotive Industry	Hybrid Service	단말/서버 연동 서비스 강화
Value Shift	VUI Paradigm Change	음성인식 기술 중심에서 서비스 중심으로 관심 이동
	UX	사용성 개선 경쟁 심화 (로컬 기업 미들웨어 업그레이드)

로, 국내 기업이 따라잡기에는 너무 크게 격차가 벌어진 상태이다. 현재 상황에서 동시에 20여 가지의 다국어 엔진을 지원할 수 있는 능력을 갖추기란 쉽지 않은 일이다.

그러나 차량용 음성 시장에 희망이 보이는 것은 최근 일어나고 있는 음성 시장의 변화 때문이다. 엔진의 성능보다 서비스의 품질을 우선시하고, 사용자 경험에 기반한 UX 기술의 강화가 요구되면서 음성 미들웨어를 통한 경쟁력 확보가 가능한 시점에 이르렀다. 따라서 다국어를 지원하면서 동시에 다양한 사용자의 요구 사항을 만족할 수 있는 음성 미들웨어의 개발을 통해 점차 커지고 있는 음성 시장에 참여하고, 해외 진출을 통해 국내 음성 산업 경쟁력을 높이는 것이 가능

하다.

### 3. 다국어 음성 미들웨어와 언어 현지화

차량용 음성인식 시스템은 우수한 엔진을 보유하고 있다고 해서 상용화에 반드시 유리한 것은 아니다. 음성인식 엔진을 활용하여 자동차에서 요구되는 기능을 원활하게 수행할 수 있는 능력과 기술을 보유했는가 하는 점이 보다 더 중요하다. 자동차 설계 요구 사항을 정확히 구현할 수 있는지, 일정을 준수할 수 있는지 등의 요인을 만족시킬 수 있어야 상용화 성공률을 높일 수 있다. 차량 환경에 음성 미들웨어를 도입하는 이유는 다른 분야에 비해 복잡한 차량용 멀티미디어 전

장 기기들 사이의 연동을 조화롭게 수행하기 위해서이다.

일단 상용화가 진행되면 사양의 변경 및 소비자 요구 조건 수용 등의 문제로 이미 구현된 소프트웨어의 기능을 변경하거나 새로운 기능을 추가해야 하는 상황이 자주 발생한다. 지속적인 개선과 발전을 통해 국산 신차의 성능과 가치를 향상시켜야 하는 시점에서 비용의 상승과 일정 지연은 큰 결점으로 작용할 가능성이 높으므로 개선 방안이 요구된다.

대체로 상품성 및 성능과 관련된 민감한 부분의 수정은 긴급한 처리가 필요한 경우가 대부분이다. 이러한 상황은 소프트웨어의 변경을 최소화하여 대응하는 전략이 필요하며, 이를 위해 음성 엔진과 음성 미들웨어, 장치 연동 어플리케이션의 독립적 모듈화가 큰 효과를 발휘한다.

### 3.1 데이터 기반 차량용 모듈형 음성 미들웨어

음성 미들웨어를 음성 엔진과 분리된 독립적인 모듈로 구성하면 여러 회사의 음성인식 및 음성 합성 엔진을 도입하더라도 쉽게 사용할 수 있는 호환성을 확보할 수 있다. 이때 음성 미들웨어에서 제공하는 표준화된 엔진 API 사용을 통해 엔진 교체에 따르는 비용 부담 및 소요 일정을 경감시킬 수 있다.

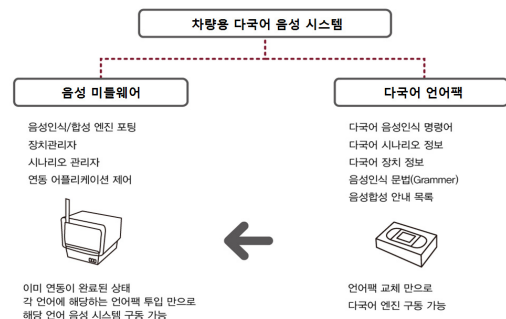
데이터 기반 방식의 모듈형 음성 미들웨어는 음성인식 및 합성 엔진의 포팅, 장치 관리, 시나리오 관리, 연동 어플리케이션 제어를 담당하는 소프트웨어 플랫폼 부분과 다국어 음성 명령어 데이터, 시나리오 데이터, 음성인식 문법(Grammar) 정보 등을 저장하고 있는 콘텐츠 부분인 다국어 언어팩(Language Pack)으로 나누어 볼 수 있다.

(그림 2)에서와 같이 모듈형 음성 미들웨어는

하부 어플리케이션과의 프로토콜만 정의되면 내비게이션 및 멀티미디어 장치와 쉽게 연동할 수 있는 프로토콜 전달 구조를 획득할 수 있다. 이러한 방식을 통해 엔진과 미들웨어, 연동 어플리케이션 사이의 정보를 양방향 통신 방식으로 관리할 수 있는데, 이는 실시간으로 엔진 및 어플리케이션의 상태를 파악하고 제어하는 데에 매우 유리한 구조를 제공하며, 이를 통해 음성인터페이스의 원활한 흐름(flow)을 보장받을 수 있게 된다.

이처럼 독립적으로 모듈화된 구조를 바탕으로 하여, 하위 계층의 문제가 전체 시스템에 영향을 준다면, 한 부분의 문제가 다른 부분에 영향을 주는 경우, 음성인식 및 합성 엔진 교체 시 전체 프로세스의 변경을 수반하는 경우 등의 단일 프로세스에서 발생하는 문제를 야기하지 않는 안정적인 서비스 제공을 가능하게 한다. 이러한 구조의 장점은 기본적인 프로그램 동작의 틀은 그대로 둔 채, 엔진과 콘텐츠의 변경만으로 다양한 언어의 음성 시스템을 간편하게 구현할 수 있다는 점이다.

다국어 콘텐츠를 저장하고 있는 언어팩에는 기본적인 명령어, 시나리오 흐름, 기능 및 장치 제어 등의 콘텐츠가 포함되어 있다. 전체 VUI 설계 사양의 대부분을 언어팩에서 구현하며, 어플리케이션의 상태에 따른 다이얼로그 천이 정



(그림 2) 차량용 음성 미들웨어 구조

보 및 사용자 인터페이스 정보들을 포함한다. 이러한 데이터는 각 언어별로 별도 작성되며, 각 해당 언어의 언어적 특성, 지역적 특성, 문화적 특성 등이 고려된 현지화(Localization) 작업을 거쳐 실제 양산 후에도 편리한 사용성을 확보하도록 구현된다.

이러한 언어팩의 데이터 기반(Data Driven) 방식은 자동차 환경의 음성인식 시스템 양산 과정에서 몇 가지 우수한 효과를 불러올 수 있다.

우선 첫 번째로 신속한 명령어 교체가 가능하다. 사양이 최종적으로 결정되지 않은 상태에서도 초기 버전 구현이 가능하고, 추후에 명령어의 변경이 있다 하더라도, 하드코딩 없이 데이터 부분만 수정하면 되므로 즉시 명령어 교체가 가능하다. 이러한 특성은 다른 다국어의 경우에서도 모두 동일하게 적용된다.

두 번째로 편리한 시나리오 변경이 가능하다. 음성인식의 수행 과정은 여러 가지의 분기와 선택, 판단의 과정에 내재되어 있기 때문에 사양의 변경이 있을 경우 이를 일일이 하드코딩으로 구현하기에는 예외 처리의 경우가 너무 많이 생기게 된다. 시스템의 복잡도 증가는 필연적으로 수정 및 대응의 지연과 비용의 증가를 수반하기 때문에 빠른 양산 대응에는 매우 불리한 구조이다. 따라서 데이터 기반(Data Driven) 방식을 이용한 언어팩 개발은 문제점 수정 및 사양 변경에 매우 유리하며, 초급 개발자도 툴을 이용하여 무리 없이 데이터 생성 및 수정에 참여할 수 있으므로 접근성이 뛰어나다.

세 번째로 간편한 기능 및 장치 적용이 가능하다. 언어팩에서는 각 기능에 대한 여러 명령어를 설정하도록 하는데, 차종별로 서로 다른 기능 및 장치가 적용되는 경우 데이터 기반(Data Driven) 방식을 통해 이를 쉽게 관리할 수 있는 장점이 있다.

네 번째로 해당 데이터에 대한 재가용성이 매우 높다는 것이다. 예를 들어 기존의 언어팩 데이터를 새로운 플랫폼에 적용할 때, VUI 사양 가이드라인이 크게 바뀌지 않는다면, 상당 부분의 데이터를 재가용할 수 있다. 이는 곧 새로운 플랫폼 개발 시 소요되는 개발비를 크게 줄일 수 있는 장점으로 작용한다.

### 3.2 다국어를 위한 언어 현지화 (Localization)

최근 국내 자동차 업계의 발전으로 수출 물량이 크게 늘면서 다국어 음성인식에 대한 수요가 급증하는 추세에 있다. 이처럼 점차 높아지는 다국어 수요에 맞추어 다국어 대화 처리가 가능한 차량용 음성 미들웨어 및 대화처리 엔진의 개발이 요구된다. 각 언어마다 서로 다른 명령어, 상이한 주소 체계, 전화번호 체계, 전화걸기 이름 등 각 언어에 맞는 데이터를 모두 가공하여 시스템에 보유하고 있어야 한다. 따라서 다국어 음성인식을 위해서는 해당 언어에 대한 정확한 표현을 반영해야 하는데, 이를 위해 진행되는 작업이 현지화(Localization) 작업이다.

소극적 의미에서 언어 현지화는 단순히 번역을 뜻하기도 한다. 그러나 적극적인 의미에서 언어 현지화는 외국에서 만들어진 콘텐츠를 현지에서 아무런 불편 없이 사용할 수 있도록 하는 것을 의미한다. 예를 들어서 “Call James”라는 영어식 명령어는 한국어로 번역할 경우 “전화걸기 제임스”로 번역할 수 있다. 물론 이러한 명령어를 그대로 사용하는 것도 가능하지만, 우리의 입장에서는 “제임스에게 전화걸기”가 더 자연스럽다. 음성인식의 명령어를 잘 숙지하지 못한 일반 사용자의 경우, 영어식 명령어인 “전화걸기 제임스”보다는 한국어식 어순인 “제임스에게 전화걸기”라고 명

령어 발화를 할 가능성이 높다. 이처럼 언어마다 다른 어순을 고려하여 자연성을 높이기 위한 노력을 한다거나 언어마다 차이가 생기는 요소들을 적절하게 조정하는 역할을 하는 작업이 바로 언어 현지화 작업이다.

예를 들어 주소검색 기능의 경우, 각각의 주소 체계는 나라마다 다르므로, 이 체계에 맞게 음성 인식 문법(Grammar)을 생성하여야 한다. 이때 주소 입력을 위한 번지의 경우 각 언어마다 숫자 읽는 방식의 차이에 따른 변경 요소들을 잘 반영해야 한다. 전화걸기 서비스의 경우 보통 성(First Name)과 이름>Last Name)을 구분하여 인식 문법을 생성해야 할 경우가 있다. 한국과 달리 성과 이름의 순서가 바뀐 경우가 있으므로, 이러한 부분도 반영이 되어야 한다. 또한 전화번호의 경우 나라마다 다른 번호 체계가 고려된 번역 결과를 반영해야 하며, 전화를 걸기 위한 명령어와 번호의 순서 등을 현지 사정에 맞게 잘 반영해야 한다. 라디오 주파수의 경우 주파수 사용 범위가 다르고, 한국

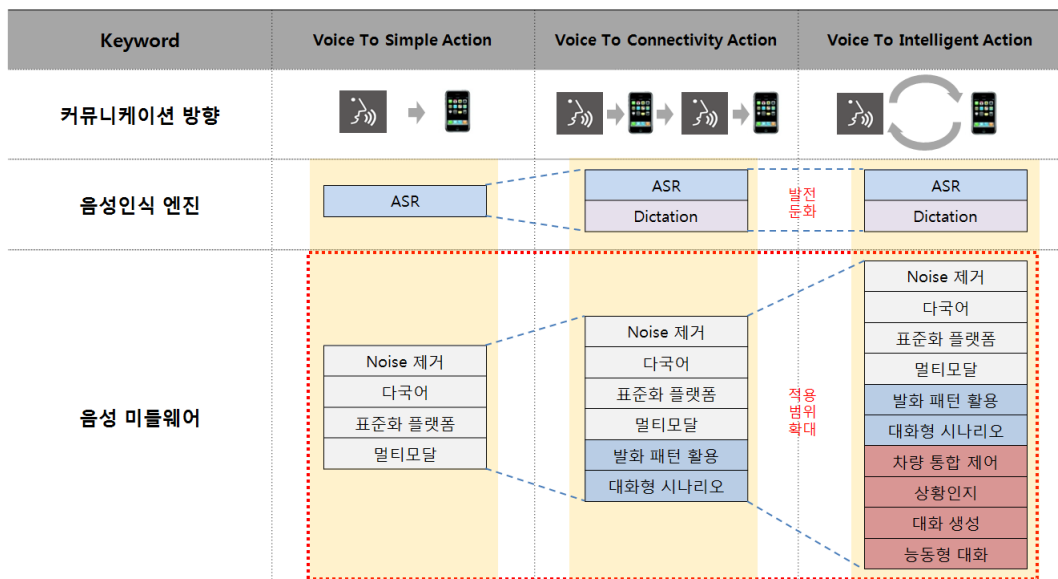
에는 없는 XM 위성 방송이나, 판도라 등의 방송국명 등 그 지역 환경에 맞는 언어 변환이 필요하다.

이렇게 각 언어에 맞는 언어 현지화의 작업은 단순한 번역을 넘어서는 정밀한 조사 및 검토 작업이 수반되어야 하는 중요한 작업이다.

#### 4. 차량용 음성 미들웨어 발전 방향

최근 서버 기반의 딥러닝 엔진이 속속 상용화되면서 음성인식 엔진 분야는 거의 최고 기술 수준에 이르렀다고 볼 수 있다. 이제부터는 딥러닝 결과를 어떻게 서비스로 연결하는가의 문제가 중요한 이슈로 등장하고 있다.

(그림 3)과 같이 음성인식 엔진분야보다는 음성 미들웨어 분야의 발전 가능성이 더 높은 상태이며, 상황인지 및 능동형 대화 시스템과 같이 지능적인 서비스를 구현하기 위해서는 음성인식 엔진의 결과를 이용하는 후처리 모듈의 중요성이 더 강조된다. 음성 미들웨어에서 이러한 후처리



(그림 3) 차량용 음성 미들웨어 발전 방향

를 담당하게 되면, 차량 내 연동 기능과 음성인식 엔진 포팅 부분의 변경을 최소화하면서 신기술의 적용 및 새로운 서비스 개발에 효과적으로 인력과 시간을 집중할 수 있다.

본 장에서는 기존의 차량용 음성 인터페이스 기술의 여러 한계를 극복하기 위해 보완되어야 할 기술 항목에 대해 간략히 살펴보고자 한다.

#### 4.1 발화 자연성 증대

음성인식 기술이 가진 여러 장점에도 불구하고, 사용성 측면에서의 문제 제기가 계속되어 왔다. 그 중에서도 일상생활의 대화에서 자주 나타나는 무의미한 감탄사나 머뭇거림 등이 명령어에 포함될 때 인식 성능이 저하되는 문제는 발화 자연성 증대의 관점에서 개선 노력이 필요하다. 발화 자연성 증대 기술의 핵심은 간투사 제거 기술과 문장형 인식 기술의 적용이다.

간투사 제거 기술은 사용자가 편하게 음성 명령어를 발화할 수 있도록 실제 대화 환경과 유사한 처리를 하는 기술을 말한다. 사용자의 발화 실수나 머뭇거림, “음...”, “아...” 등의 간투사 처리 등을 통해 실제 대화 환경에서 일어날 수 있는 돌발적인 상황에도 음성인식 기능이 정상적으로 처리될 수 있도록 한다. 기존의 음성 시스템은 이러한 머뭇거림이 입력되었을 때 대개 오인식 처리되어 사용자의 감성 품질을 크게 저하시키는 원인이 되어 왔다.

한편 발화자의 의도를 파악하여 원하는 기능을 수행하도록 하는 가변 문장 형식의 명령어 입력은 소비자 사용성을 높일 수 있는 좋은 방안이 될 수 있다.

지금까지 차량에서 양산된 대부분의 음성 기술은 미리 정해진 고립형 단어 위주의 고정된 명령어 발생 패턴을 인식하는데, 이는 주행 상황에서

정확한 인식 성능을 보장하기 위해서였다. 그러나 이러한 방식은 매우 치명적인 단점을 가지고 있는데, 그것은 바로 사용성이 너무 떨어져서 사용자에게 심각한 불편함을 야기한다는 점이다. 명령어를 잘 모르는 경우 정해지지 않은 평소의 표현을 사용하게 되는데, 이러한 경우 대부분 오인식이 발생하게 된다. 이는 지금까지 음성인식에 대한 부정적 인식과 기술에 대한 신뢰도 하락 현상이 발생하게 된 주요 원인이 되어 왔다.

예를 들어, “내비게이션”이라는 명령어를 정확하게 발생해야 인식이 되는 시스템에서 평소와 같이 “내비 켜봐라, 내비 켜줘, 내비게이션 켜줄래?, 지도 보여주세요” 등의 자연스러운 대화형 문장을 입력해도 내비게이션이 동작하게 함으로써 소비자의 만족도를 향상시킬 수 있다.

문장형 데이터의 수집과 통계적 처리를 통한 후처리 모듈을 음성 미들웨어에 추가함으로써, 이러한 대화형 명령어 입력이 가능하다. 이러한 기술은 소비자가 원하는 의도를 파악하고 이를 대화형식으로 응답하는 지능형 대화 시스템으로의 발전이 가능하다.

#### 4.2. 연결성(Connectivity)

텔레매틱스 서비스와 인포테인먼트 검색 등을 가능하게 하는 무선 인터넷의 차량 적용은 운전자들에게 보다 유용한 정보를 제공할 수 있게 한다. 이러한 연결성의 확보를 통해 서버 기반의 음성 시스템을 사용할 수 있으며, 대용량 음성 검색 엔진 및 딥러닝 엔진 등을 차량에서 사용할 수 있게 된다.

서버 기반 엔진을 차량에서 이용한다는 것은 매우 큰 의미를 지닌다. 지금까지 차량에 장착되는 내비게이션이나 오디오의 경우 저사양의 CPU와 메모리 등을 채택하여, 음성 시스템도 소형화

된 임베디드형 시스템만이 차량에 적용될 수 있었다. 소형화된 시스템의 한계는 성능적인 측면과 처리 속도에서 소비자의 불만을 발생시키는 주된 원인이었다. 그러나 서버 기반 음성인식 엔진을 차량에서 사용하게 됨으로써 외부 콘텐츠를 사용할 수 있음과 동시에, 대규모의 콘텐츠를 검색하기 위한 음성 인터페이스의 중요성도 더욱 부각될 수 있게 된다. 문장 형식으로 결과가 표시되는 딕테이션 엔진을 사용할 수 있게 되므로, 단어 매칭 수준이 아니라 실제 문장 분석을 통한 대화 처리의 수요가 발생한다.

특히 서버 기반 음성 검색 부분은 소비자가 원하는 의도를 파악해서 추가적인 정보를 줄 수 있도록 잘 구성된 시나리오를 제공해야 한다. 이때 기존의 방식은 미리 정해진 시나리오대로 분기하여 최종 목적 행동을 수행하는 것으로 시나리오가 마무리되었다. 하지만 서버 기반 딕테이션(Dictation) 엔진을 통해 얻는 문장 형식의 결과는 그 자체만으로 정해진 시나리오를 수행하기에 어려움이 있다. 왜냐하면 고립단어 인식 결과와는 다르게, 딕테이션 엔진의 결과는 핵심 단어들이 모두 맞더라도 사소한 몇 부분에서 오류가 발생하면 일반적인 인식 결과 텍스트 매칭을 통한 시나리오 분기 방식을 쓸 수 없기 때문이다.

이를 위해 도입하는 방식이 통계 기반의 대화 처리 모듈이다. 다양한 발화 패턴을 수집하여 모델링한 통계 기반 방식을 사용함으로써 다소 부정확한 인식 결과로도 소비자의 의도를 파악할 수 있으며, 추가적으로 더 필요한 정보는 대화를 유도함으로써 자연스러운 인터페이스를 수행할 수 있다.

### 4.3 지능형 대화 처리

자동차 운전자의 경험(UX)을 최대한 반영한

인터페이스로 대화 처리 시스템의 적용 필요성이 대두되면서, 소비자의 기대는 단말 내에서의 기기 동작만이 아니라 정보의 검색 및 메시지의 전달 등 고차원적인 수준까지 확장되었다. 차량용 음성 인터페이스의 이러한 변화는 고립단어 인식에서 문장형 입력으로, 그리고 최종적으로는 대화 처리의 단계까지 기술이 발전하게 된 원동력이 되었다.

이때 대화처리 기술은 단순한 동작(Voice to Action)의 단계를 넘어서서, 사용자 의도를 파악하고(Intent Understanding) 대화를 생성하는(Dialogue Producing) 단계까지 인터페이스 영역을 확장함으로써 사용자에게 보다 편하고 친숙한 정보 교류의 기반을 마련한다.

이러한 대화 처리 기술을 차량에 적용함으로써, 운전자가 시선을 화면으로 옮기거나 운전대를 잡았던 손을 놓고 터치를 조작하는 등의 위험한 행동을 예방할 수 있다. 또한 터치 키보드를 이용한 느린 입력 방식의 한계를 뛰어넘어서 음성만으로 빠르고 정확한 정보 검색 방법을 제공함으로써 차량 내 정보 접근성을 크게 높일 수 있다.

모바일 스마트폰과 차량의 가장 큰 차이는 운전 중에 양 손이 자유롭지 못하다는 점에 있다. 모바일 스마트폰에서 사용되는 음성인식 기술이 대체로 웹 서치 등 1차원적인 검색에 의존하고 있다는 점을 감안하면, 검색 후의 모든 행동은 사용자가 담당을 해야 한다. 하지만 차량 환경에서는 눈으로 디스플레이를 보고 터치를 할 수 없는 환경이기 때문에, 1차원적인 검색에만 의존할 수가 없다. 이러한 차이는 모바일 환경과 차량 환경이 전혀 다른 방식의 접근법을 필요로 한다는 점을 강하게 시사하는 것이라고 할 수 있다.

차량 환경은 무엇보다 음성인식과 대화처리를 이용하여, 일단 시작하면 소비자가 원하는 최종 행동(Action)까지 이루어져야 한다는 점이 중요



하다. 예를 들어, “주가 정보 알려줘”라는 명령이 있다면 모바일 환경에서는 코스피 지수, 다우 지수 등 선택할 수 있는 리스트를 펼쳐 보이며, 소비자로 하여금 추가적인 선택 및 판단을 하도록 한다. 그러나 차량 환경의 경우에는 “어떤 종목을 알려드릴까요?” 라는 질문을 통해 “한국기업” 등의 답변을 유도함으로써 최종적으로 사용자가 실제로 원하는 행동을 수행하도록 유도하는 것이 효과적이다.

음성 대화 처리 기술은 차량 환경에서 운전자가 원하는 최종의 행동을 할 수 있도록 하는 매우 중요한 역할을 담당한다. 서버 연결을 통한 다양한 정보 검색과 엔터테인먼트 영위를 통해 다양한 탐색 활동이 차량 안에서 이루어지게 되는데, 이러한 검색 결과를 어떻게 운전자에게 제공할 것인가의 문제가 본 기술이 필요한 가장 중요한 이유이다.

앞서 살펴본 바와 같이 연결성의 확보로 차량 내에서 다양한 콘텐츠의 검색 수요가 크게 늘 것으로 전망된다. 이때 자유발화에 가까운 소비자의 발화 문장을 분석하고 의도를 파악하기 위한 추가 질문을 수행하게 되는데, 이러한 일련의 과정은 스마트카와 운전자 사이에 지능적인 대화 인터페이스를 가능하게 한다.

## 5. 결론

최근 자동차는 IT 융합 기술을 적극적으로 적용하는 세계적인 흐름 속에서 예전에 비해 훨씬 더 고도화된 첨단 기술의 집약체로 거듭 태어나고 있다. 안전성에 대한 확실한 보장과 다양한 기기를 쉽고 간편하게 조작할 수 있는 편리한 기술에 대한 필요성이 대두되는 가운데, 이 두 가지 중요한 요소를 모두 만족시킬 수 있는 음성 인터

페이스 기술이 주목을 받고 있다.

차량용 음성인식 기술은 그간 성능의 문제 때문에 차량 내 적용이 어렵다는 평가를 받아왔지만 최근 차량 내 고속 환경에서도 안정적인 인식 성능을 보이는 것으로 확인이 되면서, 적용 차종이 급속히 늘고 있다. 또한 국산 자동차의 수출 급증에 따라 한국어뿐만 아니라, 다국어의 수요도 매우 높아지고 있다.

차량용 음성 미들웨어는 다양성과 이질성을 가진 음성 정보 기술 요소들을 효과적으로 통합함으로써, 사용자와 응용 개발자들에게 시스템의 내부 구현에 편리한 인터페이스를 제공할 수 있다.

이처럼 음성 미들웨어는 자동차 산업기술, 음성 처리 기술 및 정보 서비스 기술의 융합 기술로 자동차 산업뿐만 아니라 가전 산업, 정보 통신 산업, 정보 처리 산업 등과 같이 다양한 산업 분야의 전반에 걸쳐 차세대 사용자 인터페이스를 위한 기반기술로 활용될 수 있다.

향후 스마트카 시대를 맞이하여 대용량 정보 검색과 원활한 대화 인터페이스의 구현을 위해 차량용 음성 미들웨어의 지속적인 업그레이드가 필요하며, 이에 대한 높은 관심과 투자가 요구된다.

## 참 고 문 헌

- [1] 전황수, “스마트카 기술 및 서비스 동향”, 전자통신동향분석, 제27권, 제1호, 2012년 2월.
- [2] 전황수 외 1명, “국내의 자동차-IT 융합 추진동향”, 전자통신동향분석, 제24권, 제2호, 2009년 4월.
- [3] 이윤근 외 2명, “음성인터페이스 기술”, 전자통신동향분석, 제20권, 제5호, 2005년 10월.
- [4] 강점자 외 4명, “신성장동력산업용 대어휘 음성인식 기술 동향 및 응용”, 전자통신동향분석, 제23권, 제1호, 2008년 2월.

- [5] 김종진 외 3명, "내장형 음성합성 기술 동향 및 사례", 전자통신동향분석, 제23권, 제1호, 2008년 2월.
- [6] 정호영 외 8명, "자연어 음성인식 기술을 이용한 음성 대화 서비스 개발동향", 전자통신동향분석, 제26권, 제5호, 2011년 10월.
- [7] James Cannan & Huosheng Hu, "Human-Machine Interaction (HMI) : A Survey", Technical Report : CES-508, University of Essex, January, 2011.
- [8] MAIX, "Automotive Voice UI Usability Study User Survey", Nuance, February, 2009.

## 저 자 약 력



송 민 규

이메일 : minks@mediazen.co.kr

- 1998년 고려대학교 국어국문학과(학사)
- 2000년 고려대학교 응용어문정보학과(석사)
- 2007년 고려대학교 국어국문학과(박사)
- 2000년~2003년 고려대학교 민족문화연구원 음성언어 정보연구실/선임연구원
- 2003년~현재 미디어젠(주) 전략기획팀장
- 관심분야: 음성인식, 음성합성, 자연어 처리